

Best Practice Guide for the Registration of Resources with da|ra. Version 1.0

Helbig, Kerstin; Hausstein, Brigitte

Veröffentlichungsversion / Published Version
Arbeitspapier / working paper

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Helbig, K., & Hausstein, B. (2014). *Best Practice Guide for the Registration of Resources with da|ra. Version 1.0.* (GESIS-Technical Reports, 2014/18). Berlin: GESIS - Leibniz-Institut für Sozialwissenschaften. <https://doi.org/10.4232/10.bpg.1.0>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

gesis
Leibniz-Institut
für Sozialwissenschaften

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Mitglied der

Leibniz-Gemeinschaft

Best Practice Guide for the Registration of Resources with da|ra

Version 1.0

DOI: 10.4232/10.bpg.1.0

Kerstin Helbig, Brigitte Hausstein

GESIS–Technical Reports 2014|18

Best Practice Guide for the Registration of Resources with da|ra

Version 1.0

DOI: 10.4232/10.bpg.1.0

Kerstin Helbig, Brigitte Hausstein

GESIS-Technical Reports

GESIS - Leibniz-Institute for the Social Sciences
da|ra - Registration Agency for Social and Economic Data
Schiffbauerdamm 19
10117 Berlin
Germany
Phone: +49 30 23 36 11 -314
Fax: +49 30 23 36 11 -310
E-Mail: brigitte.hausstein@gesis.org

DOI: 10.4232/10.bpg.1.0
ISSN: 1869-0483 (Print)
ISSN: 1869-0491 (Online)

Publisher, printing
and distribution: GESIS - Leibniz-Institute for the Social Sciences
Unter Sachsenhausen 6-8, 50667 Köln

Contents

| | | |
|------|--|----|
| 1 | da ra | 5 |
| 1.1 | The Registration Agency da ra..... | 5 |
| 1.2 | da ra DOI Registration..... | 6 |
| 2 | How does DOI registration work? | 7 |
| 2.1 | Relation da ra - DataCite | 7 |
| 2.2 | What is meant by DOI resolution?..... | 7 |
| 2.3 | Checklist for DOI registration..... | 8 |
| 2.4 | Publication agent responsibilities..... | 8 |
| 2.5 | da ra responsibilities | 8 |
| 3 | What is a Digital Object Identifier?..... | 9 |
| 3.1 | Examples of da ra suffixes | 10 |
| 4 | da ra Metadata Schema..... | 11 |
| 4.1 | Mandatory da ra metadata properties..... | 11 |
| 5 | Citation of resources..... | 12 |
| 5.1 | Examples for resource citations..... | 13 |
| 6 | Versioning..... | 14 |
| 6.1 | Examples for the versioning of resources | 15 |
| 7 | Granularity | 16 |
| 7.1 | Examples for granularity | 17 |
| 8 | Assurance of data quality..... | 18 |
| 9 | Assurance of metadata quality..... | 19 |
| 9.1 | High quality metadata examples from da ra | 20 |
| 10 | Linking of resources | 21 |
| 10.1 | Example for a related resource via related identifier..... | 22 |
| 11 | Glossary | 23 |
| 12 | Resources and references | 24 |
| 12.1 | da ra | 24 |
| 12.2 | DataCite..... | 24 |
| 12.3 | International DOI Foundation | 24 |
| 12.4 | Other references..... | 25 |

1 da|ra¹

1.1 The Registration Agency da|ra

da|ra operates as the registration agency for social and economic data jointly run by the GESIS – Leibniz Institute for the Social Sciences (<http://www.gesis.org>) and ZBW Leibniz Information Center for Economics (<http://www.zbw.eu>). da|ra pursues the goal of long-term, persistent identification and availability of research data via allocation of DOI names. In keeping with the ideals of good scientific practice there is a demand for open access to existing primary data so as to not only have the final research results but also be able to reconstruct the entire research process. GESIS and ZBW therefore offer a registration service for social and economic research data in cooperation with DataCite (<http://www.datacite.org>), an international consortium pursuing the goal of supporting the acceptance of research data as independent citable scientific objects. This infrastructure lays the foundation for long-term, persistent identification, storage, localization and reliable citation of research data.

Benefits of DOI names:

- **Permanent, persistent identification:** Each DOI name uniquely, unequivocally and permanently identifies the assigned object.
- **Availability of information on the web:** Via the Handle System, each DOI name refers to one or more webpages assigned by the publication agent.
- **Semantic Interoperability:** The metadata associated with a DOI name enables direct, precise communicating – with each user, from every location, at every point in the production/distribution chain – with regard to every detail of the objects related with one another.

The DOI name is comprised of a unique alphanumeric character string; a prefix and suffix, whereby the prefix always begins with "10" and prefix and suffix are separated by a forward slash. Prefixes are assigned by the International DOI Foundation (IDF <http://www.doi.org>) via DataCite. Each data centre is assigned its own prefix thus permitting an unlimited number of DOI names. The suffix is agreed by the publication agent in conjunction with da|ra.

Each DOI name permanently identifies the assigned object as an entity regardless of whether the storage location changes. Updated, structured metadata is assigned to the resource using the DOI name. The allocation of DOI names to the objects transpires automatically following successful transmission of the metadata per object to be registered.

In 2013, da|ra began to develop value-added services. These included preparations for the extension of the registration service to other resource types. As social science and economic research produces not only datasets, but also other resource types it was a logical step. da|ra Version 3.0 was therefore expanded for the registration of the resources Collection, Text, Image, Video, Audio and Interactive Resource.

¹ This guide was inspired by British Library (2013): Working with the British Library and DataCite. A guide for Higher Education Institutions in the UK. Chapters 2.1 – 2.3 are based on that guide.

1.2 da|ra DOI Registration

da|ra obtains the DOI names via the GESIS membership in DataCite. DataCite is accredited as an official DOI registration agency within the DOI foundation (IDF <http://www.doi.org>).

TIB Hanover (<http://www.tib.uni-hannover.de/>) acts as a managing agent of DataCite and organizes the control of prefixes and the connection to IDF. The figuration of the suffixes is done by the publication agents and is determined in the Service Level Agreement².

da|ra governs the assignment of DOI names. It functions as the DOI allocation agency and is not commercially oriented. Besides the DOI allocation, da|ra is responsible for the elaboration of the service agreement (Service Level Agreement) together with the publication agents as well as for the administration of metadata. The data centres are responsible for both the maintenance and the storage of metadata.

Ensuring that metadata is persistent does not exclude its modifiability: data producers have the opportunity to amend the metadata whenever and as often as needed.

² da|ra [Service Level Agreement](#) (February 2014)

2 How does DOI registration work?

2.1 Relation da|ra – DataCite

You can become a publication agent with the Registration Agency for Social and Economic Data da|ra. da|ra is a member of DataCite and sends a part of the delivered metadata to DataCite. DataCite is a member of the International DOI Foundation (IDF) and one of a handful of registration agencies (e. g. also CrossRef) which have the authority to allocate DOI names. It is the only agency that focuses on identifiers for research data. da|ra has the authority to allocate DataCite DOI prefixes to its publication agents.

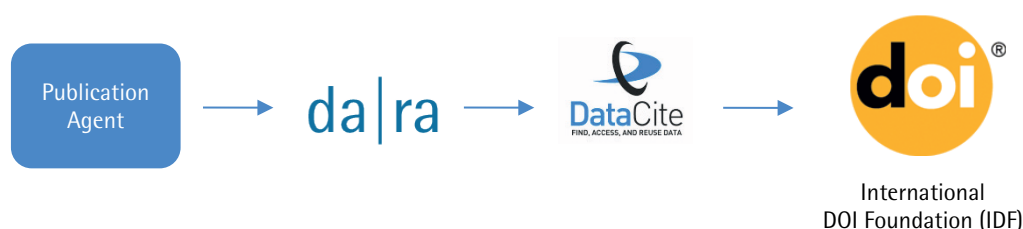


Figure 1: Metadata delivery

2.2 What is meant by DOI resolution?

DOI resolution is made possible by the Handle system technology developed by CNRI (<http://www.cnri.reston.va.us/>). Simply enter the text string doi.org, followed by the DOI name in the address bar of a preferred browser.

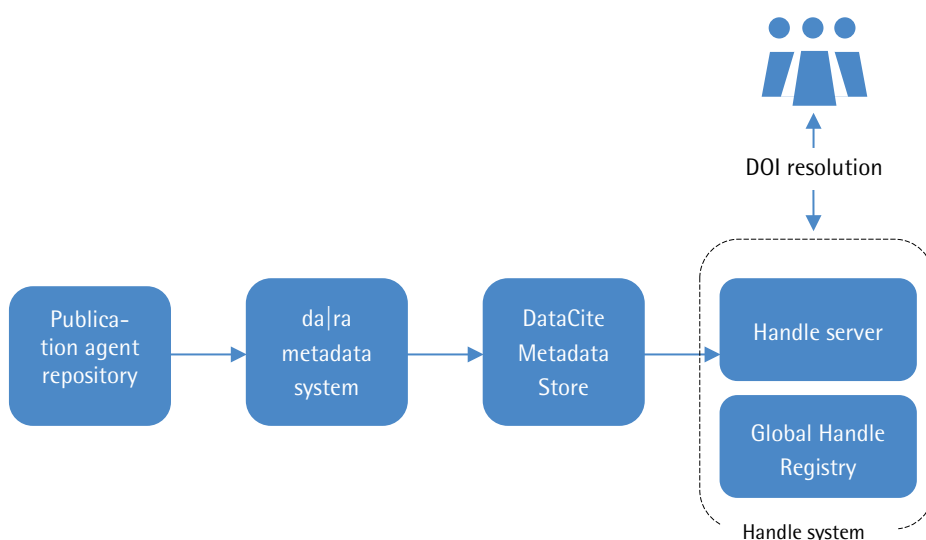


Figure 2: DOI resolution

An example: A DOI name of the form 10.4232/1.10059 would be resolved from the address: "http://doi.org/10.4232/1.10059".

To make it easier for users to access an object linked with a DOI name, DataCite as well as da|ra offer a so-called Resolver (Handle System Proxy) which offers fast and easy resolution of a DOI name.

A DOI name resolves to a landing page where metadata information is stored. Access to the dataset will be determined by the present conditions of the storage location. The intended goal is to gain direct access to the dataset with the next click. Within DataCite it is Best Practice to resolve a DOI name to a corresponding landing page.

2.3 Checklist for DOI registration

✓ **You have the authority to assign DOI names to your data**

If you do not own the data you must have permission from the data owner/creator to assign DOI names.

✓ **You can guarantee data persistence**

Before signing a Service Level Agreement³ with da|ra you might be asked to demonstrate a commitment to maintaining your data in an accessible and usable form for the long term.

✓ **The data is accessible to external users**

The central aim of DataCite is to make research data citable, therefore DOI names are not suitable for data that is completely unavailable to users outside your institution.

We understand that, for certain types of data, it may be necessary to impose access restrictions or embargoes. In such cases, information about how to obtain and use the data should be made clear at the point of access (i. e. the landing page).

✓ **The data has citation potential**

In other words, it is likely to be of interest to other researchers/users and may be cited in future works.

2.4 Publication agent responsibilities

Potential publication agents are organizations, or units thereof, that can demonstrably fulfill the requirements for digital objects, metadata and storage systems listed in the currently valid [da|ra Policy](#). To guarantee permanent and reliable data accessibility in the sense of the concept of the persistent identifier is the foremost priority.

2.5 da|ra responsibilities

The non-profit organization da|ra acts as the issuing agent for assigning DOI names in accordance with the principles of DataCite. da|ra is operated by GESIS (for social sciences data) and ZBW (for economic data) and is responsible for managing the related service level agreements with publication agents and administering metadata.

³ da|ra [Service Level Agreement](#) (February 2014)

3 What is a Digital Object Identifier?

A Digital Object Identifier (DOI) is a permanent, persistent identifier used for citing and linking electronic resources (texts, research data or other content). Updated, structured metadata is assigned to the document using the DOI name.

A DOI is not case sensitive, i. e. there is no distinction between capital and lowercase letters in a text string (for example, 10.1392/roma081203 is the same as 10.1392/ROMA081203). This identifier is purely a simple character string and outside of the DOI system, nothing can be derived from it.

The DOI name is comprised of a unique alphanumeric character sequence. The syntax of a DOI name consists of a **prefix** and a **suffix**. Both are separated by a forward slash („ / ").

10.1234/5678abcd
prefix suffix

Prefixes are assigned by the DOI Foundation via DataCite/da|ra. The suffix is agreed on by the publication agent in conjunction with da|ra. An automatic generation via da|ra is possible as well. The suffix is arranged and declared in the form following the general structure conformable to § 2 of the [Service Level Agreement](#) (SLA):

- (1) XX registers objects on [study level. I.e. one DOI name can be linked with several data sets of a certain study.] [data set level] [article level] [...] Revisions of the granularity are feasible according to prior agreement of da|ra.
- (2) XX receives the DOI-prefix "10.yyyyyy"
- (3) [The suffixes have the following agreed structure: /...][The suffix will be generated automatically by XX.]
- (4) Every suffix has to be biunique within the prefix.
- (5) The domain(s) under which XX will register objects is/are: [...] [...] ...
The URLs for all landing pages have to come from this domain.
da|ra has to be informed about changes to the domain.

Furthermore please note the following **recommendations** for the design of DOI names:

The suffix should be an opaque string, which is composed of

- a sequential number or
- an internal identifier (number).

Prefix and suffix are not constrained in their length. As the DOI name can also be used as a Uniform Resource Locator (URL) only the following signs are allowed:

The letters A-Z as well as a-z, the numbers 0-9, - (hyphen), . (point), _ (underline), : (colon) and / (forward slash, but only in front of an alphanumeric character string).⁴

As an illustration some recommendable examples follow.

⁴ Not allowed are the following signs: Language-specific vowels (e. g. ä, Ä, ß), %, ", #, <, >, }, {, ^, [,], ', |, \ and spaces.

3.1 Examples of da|ra suffixes

These examples are from da|ra publication agents (PA) and show you how a suffix can be designed.

| Complete DOI name | Prefix | Suffix | PA | Comment |
|---|---------|----------------|-------|---|
| 10.5157/ NEPS:SC6:1.0.0 | 10.5157 | NEPS:SC6:1.0.0 | NEPS | Semantic part (NEPS), internal identifier (SC6) and version (1.0.0) |
| 10.5684/ soep.v27 | 10.5684 | soep.v27 | SOEP | Semantic part (soep) and version (v27) |
| 10.4232/ 1.11004 | 10.4232 | 1.11004 | GESIS | Sequential number |

4 da|ra Metadata Schema

The da|ra Metadata Schema is a list of core metadata properties chosen for the identification of data and retrieval purposes. Each DOI name is linked to a set of metadata, a collection of bibliographical and content information, which describe in detail the registered resources (title, author, publication date, copyright etc.) and present the properties of resources, their structure and contextual relations.

The da|ra Metadata Schema provides a determined number of mandatory elements – core properties –, that have to be submitted by the publication agent at the time of registration.

4.1 Mandatory da|ra metadata properties

There are six mandatory metadata properties in the da|ra Metadata Schema (currently Version 3.1; [doi:10.4232/10.mdsdoc.3.1](https://doi.org/10.4232/10.mdsdoc.3.1)) that have to be provided when you register a DOI name via da|ra:

| Property | Definition |
|---------------------------|--|
| General Resource Type | The general type of a resource (collection, dataset, text, video, image, audio or interactive resource). |
| Title | The title of a resource. |
| Creator | The name of the principal investigator or author. May be a corporate/institutional or a personal name. |
| URL | Each DOI name has an URL to which it resolves (landing page). |
| Publication Date | The publication date of the resource submitted by the Publication Agent; possible are format sub-properties date; monthyear or year. |
| Availability (controlled) | Conditions governing the access to primary resource. |

Note: Once registered, the mandatory metadata should not be changed. Please ensure that the metadata is free of errors.

Publication agents may also choose to use optional properties to identify their resources more clearly. Optional or recommended metadata properties and sub-properties may be updated or added to as required. Please also note the Best Practice recommendations "[Assurance of metadata quality](#)".

For all metadata properties the respective names, definitions, attributes, conditions, cardinality (maximum occurrence) as well as value domains are defined. Some properties comply with ISO norms. These norms determine e. g. which code for a language or geographic coverage has to be applied. Controlled vocabularies such as thesauri and classifications are applicable. These vocabularies are complemented by da|ra controlled terms.

Although da|ra complies with the official DataCite Metadata Schema, it has broadened the DataCite metadata by adding some specific properties related to the social sciences and economics. Please note that da|ra reserves the right to share metadata with information indexes and other entities.⁵ da|ra metadata are subject to [Creative Commons](#) CC0 license.

⁵ See da|ra Service Level Agreement and Policy in its respective current form <http://www.da-ra.de/en/about-us/da-ra-policy/>.

5 Citation of resources⁶

The usage and analysis of research data as well as the publication of research results require information on author, location and identification of data. On the basis of the da|ra Metadata Schema (currently Version 3.1, November 2014) da|ra recommends the following citation style for registered resources, which is also applied in its own information system:

Creator (Publication date): Title. Version. Publication agent. General resource type. Identifier

Creator: The creators, i. e. primary researchers or authors are displayed in the order in which they were added. Only the name is used for people, not the affiliation. If only an institution is specified as the creator that is used (see 5.1 example 1). The data about the creators can include up to five names, beyond that they are abbreviated with 'et. al.' (see 5.1 example 2).

Publication date: The publication date is the year in which the resource was published in digital form by the publication agent.

Title: Here, the main title of the resource and any additional title must be indicated. Additional titles include:

- a) **Other titles:** All subtitles must be stated in the citation. Main and subtitles are separated with ".". Alternative and translated as well as original titles are not quoted in the citation.
- b) **Collective title:** A collective title, if existent, is also specified in the citation (see 5.1 example 3).

Version: The version represents the version number of the resource.

Publication agent: The publication agent is the name of the institution or of the data centre, which has published the resource.

In the da|ra information system there are also entries that do not contain any information on the publication agent (entries within the da|ra reference system; see <http://www.da-ra.de/en/for-researchers/reference-data/>). In the citation generated from these entries no publication agent is listed.

General resource type: The general resource type describes the general type of the resource. The resource types include collection, dataset, text, video, image, audio and interactive resource.

Identifier: The identifier is the *Digital Object Identifier* (DOI). It is displayed in the citation as a URL link including the necessary resolver and without "doi:".

There are also entries in the da|ra information system that do not have a DOI name but only an URL (Uniform Resource Locator). (These are entries within the da|ra reference system; see <http://www.da-ra.de/en/for-researchers/reference-data/>). In these cases, the last access date follows the URL in the citation.

⁶ The recommendations are based on the instructions in the DataCite Metadata Schema Version 3.1 (November 2014) <http://doi.org/10.5438/0010>

5.1 Examples for resource citations

These citation examples describe resources of da|ra publication agents. They show you how the optimal data citation should look like and how the citation for registered resources is implemented in the da|ra information system respectively.

GESIS Panel Team (2014): GESIS Panel – Standard Edition. Version: 2.0.0. GESIS Datenarchiv. Dataset. <http://doi.org/10.4232/1.11947>

Markowski, Radosław; Gebethner, Stanisław; Grabowska, Mirosława; Grzelak, Paweł; Jasiewicz, Krzysztof et. al. (2006): Polish National Election Study 2000 (PGSW). Version: 1.0.0. GESIS Data Archive. Dataset. <http://doi.org/10.4232/1.4334>

Wagner, Michael; Valdés Cifuentes, Isabel (2014): The Pluralisation of Living Arrangements – A Continuous Trend?. Comparative Population Studies. Version: 1.0.0. CPoS – Federal Institute for Population Research (BiB). Text. <http://doi.org/10.12765/CPoS-2014-03en>

6 Versioning

The versioning is critical for the long-term preservation of resources. This chapter serves to provide you with an overview of the process of versioning and to answer first questions that may arise. If something should remain unclear, you can contact us via the da|ra contact form (<http://www.da-ra.de/en/contact/>)

In general the following aspects should be considered regarding versioning:⁷

- An object with an assigned DOI name should not be changed.
- Each change must be saved as a new version and a new DOI name must be assigned.
- The publication agent is responsible for versioning.

At the beginning, it is useful to define milestone versions of the resource and to store those separately. Changes in these milestone versions should be documented and saved as a new version (see first example below). There are various ways to formulate versioning.

However this is essential: unity and system!

Our recommendations regarding versioning (three-digit version number X.X.X):

- Increase of the first digit if new data is added (e. g. waves, samples etc.)
- Change of the second digit if corrections are made, which influence the analysis (e. g. change of values of respondents)
- If the documentation is changed or emended (typing error or more detailed text added etc.) only the third digit will be increased

As an illustration some recommendable examples follow.

⁷ From [da|ra Policy](#) Version 2.2 (March 2013)

6.1 Examples for the versioning of resources

These examples describe resources of da|ra and DataCite publication agents and show you how the versioning of objects should ideally look like.

| Resource (e. g. dataset) | Information about versioning |
|---|--|
| Schmitt-Beck, Rüdiger et al. (2009): Short-term Campaign Panel (GLES). Version 3.0.0. GESIS – Leibniz-Institute for Social Sciences. Dataset. doi:10.4232/1.11131. http://doi.org/10.4232/1.11131 | 1.0.0 was Pre-Release doi:10.4232/1.10364 2.0.0 The dataset has been supplemented with the missing weights (cross-sectional and panel weights) doi: 10.4232/1.10365 3.0.0 Incorrect values were replaced and new variables added doi:10.4232/1.11131 |
| Office for National Statistics (2012): Quarterly Labour Force Survey 1992–2011: Secure Data Service Access. 3. Edition. UK Data Archive. Dataset. doi:10.5255/UKDA-SN-6727-2. http://doi.org/10.5255/UKDA-SN-6727-2 | 1. Edition was the release 2. Edition minor changes and DOI created doi:10.5255/UKDA-SN-6727-1 3. Edition further data was added doi:10.5255/UKDA-SN-6727-2 |

7 Granularity

Granularity is an important aspect when registering research data. Granularity describes the degree of aggregation of the resource to be registered. Different levels of granularity can be useful depending on discipline or resource.

According to **da|ra Policy** this is first to note with granularity:⁸

The publication agent receives a fixed prefix from da|ra which must be used for all the objects to be registered. The identification of an object can be executed to any desired level of granularity (element of a file, a file, a file collection, etc.) as the purpose dictates. Recommendations of the IDF (<http://www.doi.org/hb.html>) should be taken into consideration. The basic granularity will be adhered to when allocating a DOI name in accordance with the SLA.

The **Service Level Agreement** provides for DOI name allocation and granularity the following:⁹

The publication agent registers objects at the [study-level. i.e. one DOI name can be linked with several data sets of a certain study.] [...] Changes in level of granularity are possible by prior agreement with da|ra.

The **International DOI Foundation (IDF)** comments on granularity as follows:¹⁰

A DOI name can be assigned to any object, regardless of the extent to which that object might be a component part of some larger entity. DOI names can be assigned at any desired degree of precision and granularity that a registrant deems to be appropriate.

When deciding which level of granularity is used to register a resource, the following aspects are crucial:

1. **Citation:** The current citation and research practices among the client's user community: what is likely to be cited?
2. **The use of data:** The needs of various stakeholders: how will funders/publishers/administrators etc. use the data?
3. **The type of resource:** for example a complex dataset may require a more granular identifier structure than a document or image file.
4. **Sustainability:** The client must be able to maintain each item with a DOI in accordance with DataCite client responsibilities.

As an illustration some examples follow.

⁸ See da|ra [Policy](#) Version 2.2 (March 2013)

⁹ See da|ra [Service Level Agreement](#) (February 2014)

¹⁰ See IDF Handbook Chapter 2.3.2 „Granularity“ <http://doi.org/10.1000/182>

7.1 Examples for granularity

| Resource type | Possible levels of granularity | Examples (fictitious) |
|---------------|---|---|
| Dataset | A study | Panel study "Test" |
| | A particular survey year/wave of this study | Survey year 2005 of the panel study "Test" |
| | A specific sample (subsample) within a survey year/wave of a study | Cohort 4 in survey year 2005 of the panel study "Test" |
| | Data type (SUF, PUF, online etc.), file format (EXCEL, SPSS, STATA, R) or the language version of a data file | The STATA-file (.dta) of cohort 4 in survey year 2005 of the panel study "Test" |
| Text | A journal | The journal "Testing" |
| | A specific issue of this magazine (e. g. language version) | The journal "Testing" in German |
| | An article within this journal | The article "The panel study „Test“" in the German version of the journal "Testing" |
| Video | A video (of an interview, participant observation etc.) | Video of interviews for panel study "Test" in survey year 2005 |
| | A video sequence in the video | Video of interviews from cohort 4 |

8 Assurance of data quality

The assurance of the data quality of registered resources is the responsibility of the publication agent. This is stated under the item "Quality assurance" in the [da|ra Policy](#):

"The publication agent is responsible for the quality of the contents of objects. The contents should fulfill the quality standards of the publication agent and be valid according to the rules of good scientific practice in the relevant scientific discipline."¹¹

It is da|ra a concern to support the publication agents in assuring the quality of the metadata. With this aim da|ra publishes among other things [Best Practice recommendations](#), validates metadata if the web form is used, uses controlled vocabularies according to ISO standard in the Metadata Schema and provides [advice](#) in various areas of the assurance of metadata quality and citation practice.

The focus however is the self-control and -regulation of the publication agents. Uniform and generally accepted standards for the assurance of data quality have not prevailed so far. Amongst others this is due to the difficulty to agree on a selection and hierarchy of quality criteria and to establish these criteria as binding (Schendera 2007: 13).¹²

A useful instrument for the self-assessment of data quality is the [Data Seal of Approval](#). It is used primarily in the English-speaking realm, but is also recommendable for the examination of German data providers. The [Data Seal of Approval \(DSA\)](#) ensures that archived data can still be found, recognized and used in the future. The board of the DSA, which is responsible for the review process and the further development, consists of members of [ICPSR](#), [Nestor](#), [UK Data Archive](#) and [DANS](#).

Further seals and norms for publication agents in the German-speaking realm are:

- The [nestor Seal](#) for Trustworthy Digital Archives
- The [DIN-norm 31644](#) "Criteria for trustworthy digital archives"
- The [ISO-norm 16363](#) "Space data and information transfer systems - Audit and certification of trustworthy digital repositories"

Ultimately, good scientific practice includes the assurance of data and metadata quality. Also one should bear in mind that a lack of the assurance of data quality can lead not only to financial losses, but can also damage the reputation of the data producer. Data quality should therefore be a concern of every producer and provider of data. [Here, quality goes before quantity!](#)

Further information and assistance can be found here:

- nestor-handbook: A small encyclopedia of digital long-term archiving (only in German)¹³
- Further topic-relevant DIN norms and ISO standards are for example EN ISO 19114:2005¹⁴ or DIN 55350-11¹⁵ of the DIN German Institute for Standardization

¹¹ da|ra-Policy, Version 2.2 from March 4th, 2013

¹² Schendera, Christian F. G. (2007): Datenqualität mit SPSS. München: Oldenbourg Verlag: 13.

¹³ Neuroth, H./OBwald, A./Scheffel, R./Strathmann, S./Huth, K. (Hrsg.) nestor-Hanbuch (Version 2.3) [urn:nbn:de:0008-2010071949](https://nbn-resolving.org/urn:nbn:de:0008-2010071949).

¹⁴ EN ISO 19114:2005 "Geographic information - Quality evaluation procedures"

¹⁵ DIN 55350-11 "Concepts for quality management"

9 Assurance of metadata quality

A **high metadata quality** is crucial, **because** this ensures that your data

- can be found, reproduced and reused for secondary analysis in terms of good scientific practice.
- can be linked with other data and data types (such as associated publications).
- can be correctly cited in your own interest.
- become visible after the forwarding by da|ra to other international portals (DataCite Metadata Store) and citation indices (e. g. Data Citation Index).

To increase the international visibility it is important to complete the metadata fields both in German and in **English**. In addition, **other criteria** of quality metadata are to be noted¹⁶:

- Accuracy, completeness, provenance, conformance to expectations, logical consistency and coherence, timeliness and accessibility of the metadata.

As the findability also significantly depends on the metadata quality it is advisable to consider the view of potential users of the research data and to reassess the provided metadata thereupon. Metadata quality is not to be confused with data quality (see also "Assurance of data quality"), the transition between the two is however fluent. Thus data can become metadata and vice versa.

The da|ra Metadata Schema (currently 3.1; [doi:10.4232/10.mdsdoc.3.1](https://doi.org/10.4232/10.mdsdoc.3.1)) is the central prerequisite for securing metadata quality as it offers the possibility to add additional information about the resources in excess of the mandatory properties. This additional information is recommended by da|ra and is also in the interest of the publication agents. The da|ra Metadata Schema is based on the Data Documentation Initiative¹⁷ (DDI) standard. A formal assessment of metadata correctness is effected automatically through the reconciliation of the submitted metadata in XML format (via API, XML-upload) with the XSD. The assessment for users of the web interface takes place in the background and extensive help texts are available.

A **core set of six mandatory fields** is the basic condition for the DOI registration. The provision of **additional optional elements** as well as their child elements is **recommended** explicitly for the above mentioned reasons. These metadata fields include:

Classification | Keywords | Contributor | Alternative Identifier | Description | Geographic Coverage

Authority files, e. g. the Integrated Authority File (GND) and further controlled vocabularies of the da|ra Metadata Schema give support to simplify and speed up the input. Please note that JEL-Classification can only be added on tertiary subcategory at the moment. Primary and secondary categories are not supported by da|ra 3.1.

¹⁶ Bruce, Thomas R./Hillmann, Diane I. (2004): The continuum of metadata quality: defining, expressing, exploiting. In: Metadata in Practice. Chicago: American Library Association.

¹⁷ Homepage of the Data Documentation Initiative <http://www.ddialliance.org>

9.1 High quality metadata examples from da|ra

European Values Study Longitudinal Data File 1981-2008 (EVS 1981-2008) (Dataset)
http://bit.ly/metadata_EVS_1981-2008

Going on Pilgrimage Online: the Representation of Shia Rituals on the Internet (Text)
http://bit.ly/metadata_Going_on_Pilgrimage_Online

Statistical Information System GeroStat (Collection)
http://bit.ly/metadata_GeroStat

xml examples can be found at <http://www.da-ra.de/en/technical-information/doi-registration/>.

In addition we offer [advice](#) on questions concerning the quality of your metadata.

10 Linking of resources

The potential of referencing via a DOI name can be comprehensively taken advantage of through the linking of resources. This results in a network of link, citation and recitation, which can offer great benefits for the traceability of research, secondary analysis and search.

To achieve a high quality of metadata it is important to provide as much information about the resource as possible. Some of this information are links to other resources.

- The naming of [related resources](#) using [related identifiers](#)¹⁸ can be such a connection (e. g. other versions, modified forms). The related resources can have certain graduations¹⁹.
- In addition [publications](#) to the resource (only primary literature) can be entered in a structured or unstructured way in da|ra and – if applicable – the persistent identifier can be specified.

Other identifiers of the resource as well as the primary researcher/author are of great importance for the completion of the information.²⁰ These make a precise identification easier and at the same time increase the traceability.

- [Further identifiers of the resource](#) can for example be the *Handle*, the *Persistent Uniform Resource Locator (PURL)* or the *Uniform Resource Name (URN)*.
- The [primary researcher \(person\)](#) can be described more accurately via an identifier like the *ORCID iD* or another persistent identifier. This avoids incorrect assignments and citation rates by overlap through name similarity.
- The [primary researcher \(institution\)](#) can for example be described and clearly identified with the identifier of the *Integrated Authority File (GND)* (also Universal Authority File; German: Gemeinsame Normdatei).

The completion of the metadata through the above mentioned information allows da|ra the visualization of the connections between the individual resources.

¹⁸ Element 31 "Relation" in the da|ra Metadata Schema <http://dx.doi.org/10.4232/10.mdsdoc.3.1>

¹⁹ See also the recommendations of the Florida Gulf Coast University Library Services <http://bit.ly/1a1KCek>

²⁰ See da|ra Metadata Schema <http://dx.doi.org/10.4232/10.mdsdoc.3.1>

10.1 Example for a related resource via related identifier

Tchernia, Jean-Francois; Puranen, Bi (2006): EVS – European Values Study 1999/2000 (release 2, May 2006) – France. Version: 1.0.0. GESIS Data Archive. Dataset. <http://doi.org/10.4232/1.11170>

| EVS - European Values Study 1999/2000 (release 2, May 2006) - France | | gesis |
|--|---|-------|
| DOI | 10.4232/1.11170 | |
| Version | 1.0.0 | |
| Principal Investigator | <ul style="list-style-type: none"> • Tchernia, Jean-Francois (Tchernia Etudes Conseil, Paris) • Puranen, Bi (Theseus Institute, Sophia Antipolis (France)) | |
| Temporal Coverage | <ul style="list-style-type: none"> • 1999 | |
| Related Identifiers | <ul style="list-style-type: none"> • Is previous version of doi: 10.4232/1.3776 • Is part of doi: 10.4232/1.11307 | |
| Publications | <ul style="list-style-type: none"> • The European Values Study: A Third Wave. Source book of the 1999/2000 European Values Study Surveys. Halman, Loek, Tilburg: EVS, WORC, Til ... see more | |

See: http://bit.ly/metadata_EVS_1999_2000

11 Glossary

| | |
|-------------------------------|--|
| DataCite | International initiative for establishing easier access to scientific research data; accredited registration agency at IDF |
| da ra | Registration agency for social science and economic data, run by GESIS and ZBW |
| DDI | Data Documentation Initiative, DDI is an open standard (metadata model) for the description of social science and economic data |
| DOI | Digital Object Identifier |
| DOI name | A (international unique) numerical string, directly identifying the digital object and linking it with current and structured metadata |
| DOI Registration Agency | Agency officially accredited as a registration agency at IDF |
| DOI Allocation Agency | Agency within the DataCite initiative, which allocates DOIs to the publication agent |
| GESIS | GESIS Leibniz Institute for the Social Sciences, initiator of the registration agency for social science and economic data (da ra) and partner of the ZBW |
| Granularity | Defined granularity of DOI suffix |
| Handle system | Application for the allocation and administration of persistent internet identifiers, invented by the Corporation for National Research Initiatives (CNRS); technical infrastructure for the allocation of DOI names |
| IDF | International DOI Foundation; founded for the development and administration of the DOI system |
| Persistent Identifier | Unique (location independent) identifier, ensuring reliable access to a digital object beyond a potential system change and over a long period of time |
| Publication Agent | Data provider, who registers data using DOI names |
| Service Level Agreement (SLA) | Agreement between da ra and the data provider (publication agent) |
| TIB Hannover | German National Library of Science and Technology University Library Hanover |
| ZBW | German National Library of Economics Leibniz Information Centre for Economics; partner of GESIS at the development of the registration agency for social science and economic data (da ra) |

12 Resources and references

12.1 da|ra

da|ra

<http://www.da-ra.de>

da|ra API documentation

<http://www.da-ra.de/en/technical-information/doi-registration/>

da|ra Policy

<http://www.da-ra.de/en/about-us/da-ra-policy/policy/>

da|ra Service Level Agreement

<http://www.da-ra.de/en/about-us/da-ra-policy/service-level-agreement/>

12.2 DataCite

DataCite

<http://www.datacite.org/>

DataCite Metadata Schema Repository

<http://schema.datacite.org/>

DataCite Metadata Search

<http://search.datacite.org/ui>

12.3 International DOI Foundation

International DOI Foundation

<http://www.doi.org/>

International DOI Foundation DOI Handbook

<http://www.doi.org/10.10000/182>

International DOI Foundation DOI Resolver

<http://dx.doi.org/>

12.4 Other references

British Library (2013): Working with the British Library and DataCite. A guide for Higher Education Institutions in the UK

http://www.bl.uk/aboutus/stratpolprog/digi/datasets/WorkingWithDataCite_2013.pdf

Creative Commons

<https://creativecommons.org/>

Data Documentation Initiative (DDI)

<http://www.ddialliance.org/>

Handle System

<http://www.handle.net>

ISO 26324:2012, Information and documentation -- Digital object identifier system

http://www.iso.org/iso/catalogue_detail?csnumber=43506